Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A brain splicing-QTL colocalization map of candidate effector genes for major depressive disorder

Bozhi Li

20241079@zcmu.edu.cn

Zhejiang Chinese Medical University https://orcid.org/0009-0000-6998-777X

Weijia Gu

Zhejiang Chinese Medical University

Xiaqiu Wu

Zhejiang Chinese Medical University

Article

Keywords:

Posted Date: August 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10366398/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors have declared there is **NO** conflict of interest to disclose

Abstract

Genetic risk for major depressive disorder (MDD) can act through gene expression or pre-mRNA alternative splicing, but effector-gene mapping has relied mainly on expression quantitative trait loci (eQTLs), leaving splicing-level effects less systematically resolved. We integrated the latest MDD GWAS (MDD2025; effective sample size ~ 1.15×10^6) with BigBrain, a human-brain cis-splicing-QTL (sQTL) resource of 10 725 RNA-seq samples. Within risk regions, colocalization identified 56 high-confidence brain splicing-effector genes (PP.H4 > 0.8; median 0.92), compared with approximately 22 using the 2018 GWAS. A matched BigBrain eQTL analysis separated expression-orthogonal associations from limited eQTL power: 29 genes had a genome-wide-significant cis-eQTL at the same locus that did not colocalize with MDD, defining an expression-orthogonal core; 23 had no genome-wide-significant cis-eQTL in the corresponding locus, and only four showed eQTL–MDD colocalization, with a concordant pattern in MetaBrain. Splicing Mendelian randomization supported a consistent cis-sQTL-to-MDD direction. Across disorders, 35 of 56 genes were shared with other psychiatric conditions, whereas only one colocalized with Alzheimer's or Parkinson's disease. Seventeen genes were small-molecule tractable, including CACNA1C, targeted by approved calcium-channel blockers, and HDAC3, engaged by approved non-selective HDAC inhibitors. Thirteen genes remained high-confidence under a conservative prior, and 38 of 56 loci were single-signal. In an exploratory analysis, splicing-effector genes showed higher AlphaGenome-predicted splicing disruption than background genes (62% vs 45%; one-sided P = 0.03). These results define MDD candidate genes whose associations are more clearly resolved through brain splicing than expression in current bulk-brain resources, thereby providing a complementary regulatory layer for functional and translational prioritization.

Introduction

Major depressive disorder (MDD) has a substantial polygenic basis, and recent genome-wide association studies have greatly expanded the number of associated loci.[1,2] The latest MDD GWAS, MDD2025 (Adams et al.), had an effective sample size of ~ 1.15×10^6 and identified hundreds of risk regions.[1] Because most associated variants are non-coding, interpretation requires molecular maps that connect genetic risk to regulatory consequences and effector genes.

Brain effector-gene mapping has relied mainly on eQTLs, transcriptome-wide association, SMR and colocalization.[3–5] Disease-associated variants can also act through pre-mRNA alternative splicing,[6,7] a process closely tied to neuronal development, synaptic function and neurological disease.[8–10] Large brain sQTL resources now show that splicing can explain disease genetics partly independently of expression.[11–13] In brain samples from individuals with MDD, Goes et al.[14] recently identified splicing-QTL genes at risk loci, including genes detectable through splicing but not expression. The remaining question is which MDD splicing-effector genes are supported by the latest GWAS and a larger brain sQTL resource.

Several related resources still do not provide a focused MDD splicing map. BigBrain included MDD in a pan-disease screen, but used the smaller 2018 GWAS (~ 135 000 cases) and did not provide an MDD-focused synthesis.[11,15] OPERA integrated multiple molecular phenotypes in MDD, with splicing as one layer among several.[16] Here, we treat splicing as the primary molecular phenotype, combine BigBrain brain sQTLs with MDD2025, and explicitly compare splicing signals with matched brain eQTL colocalization. We further assessed prior sensitivity, signal complexity, cross-disorder sharing, event annotation, AlphaGenome sequence-effect prediction and druggability. Our aim is a resource-oriented genetic integration rather than functional validation of a single mechanism.

Materials and Methods

Study design

This study integrated public GWAS summary statistics with brain QTL resources. We extracted BigBrain cis-sQTLs within MDD2025 genome-wide significant risk regions, tested sQTL-MDD colocalization, ran a matched brain-eQTL control, and annotated high-confidence genes by splicing MR, Steiger directionality, cross-disorder colocalization, event structure, AlphaGenome prediction and Open Targets tractability. Primary conclusions used the PP.H4 > 0.8 high-confidence set; PP.H4 > 0.5 defined an extended supplementary set.

GWAS data

MDD2025 (European ancestry, effective sample size ~ 1.15×10^6) defined risk regions, provided the GWAS outcome for colocalization and served as the outcome for splicing MR.[1] Cross-disorder analyses used schizophrenia (SCZ), bipolar disorder (BD), attention-deficit/hyperactivity disorder (ADHD), post-traumatic stress disorder (PTSD), anxiety disorders (ANX), Alzheimer's disease (AD) and Parkinson's disease (PD) summary statistics[17–23] (all data sources are listed in Supplementary Table 1). Odds ratios were log-transformed, and Z-score-only results were converted to standard errors where needed. Cross-disorder results were used for relative sharing classes, not for direct comparison of absolute effect sizes.

Brain splicing-QTL and expression-eQTL data

Brain sQTLs came from the BigBrain European-ancestry cis-sQTL resource, based on 10 725 brain RNA-seq samples from 4 656 individuals.[11] Splicing phenotypes were LeafCutter intron clusters,[24] integrated across brain sources by mmQTL random-effects meta-analysis.[25] The primary expression-level control used BigBrain brain eQTLs from the same cohort and pipeline. MetaBrain four-region eQTLs provided cross-resource validation.[26] Because BigBrain is bulk tissue, the cell-type origin of the splicing signal

was not resolved. For descriptive context only, gene-level expression across major cortical cell classes was profiled separately from single-nucleus RNA-seq (Methods; Fig. 5d).

Candidate locus definition and sQTL extraction

Genome-wide significant MDD2025 SNPs ($P < 5\times10^{-8}$) were merged into risk regions when adjacent significant SNPs lay within 1 Mb, then extended by ± 500 kb. This produced 218 regions. All BigBrain cis-sQTL associations within these regions were extracted, allowing multiple events per locus while restricting inference to MDD-significant regions.

sQTL-MDD colocalization

For each splicing event, we tested whether the cis-sQTL and MDD2025 signals shared a single causal variant using the approximate Bayes factor (ABF) colocalization framework.[3] This used a Python implementation with the standard coloc priors ($p1 = p2 = 1\times10^{-4}$, $p12 = 1\times10^{-5}$) and effect-size prior standard deviations of 0.2 for the case-control GWAS and 0.15 for the sQTL. On representative loci, the resulting PP.H4 values were identical to those from the coloc R package (coloc.abf), confirming the implementation. Because the ABF depends on the squared Z score, no allele-orientation harmonization was required. SNPs were matched by rsID, and at least 20 shared SNPs were required per test. For genes whose events spanned more than 2 Mb, the densest 1 Mb event window was used to reduce mixed multi-locus signals. Gene-level evidence was summarized by the maximum event-level PP.H4, with the number of supporting events retained. High-confidence genes were defined at PP.H4 > 0.8; PP.H4 > 0.5 defined the extended set. Where a splicing cluster mapped to multiple gene symbols, the first HGNC symbol was used consistently, with full multi-gene information retained in table notes.

Expression-level control analysis

To test whether sQTL colocalization was captured by expression, BigBrain brain eQTLs were analyzed against MDD2025 using the same regions and coloc.abf pipeline. eQTL colocalization was scored at PP.H4 > 0.8 and at a lenient PP.H4 > 0.5. MetaBrain eQTLs provided an independent check. To separate expression-orthogonal associations from limited eQTL power, each high-confidence gene was assigned to one of three tiers from the same-locus BigBrain cis-eQTLs: (i) eQTL-shared, a cis-eQTL colocalizing with MDD (eQTL PP.H4 > 0.5); (ii) expression-orthogonal core, a genome-wide-significant cis-eQTL ($P < 5\times10^{-8}$) in the locus that did not colocalize with MDD (eQTL PP.H4 < 0.5); and (iii) no/weak eQTL, no genome-wide-significant cis-eQTL in the locus. This gave 4 eQTL-shared, 29 orthogonal-core and 23 no/weak-eQTL genes. Lack of eQTL colocalization was interpreted relative to these bulk-brain resources and did not exclude expression regulation in other contexts.

Splicing MR and Steiger directionality

For each high-confidence gene with a usable instrument (45 of 56), the lead sQTL of the best-colocalizing event served as a single-instrument Wald-ratio instrument after allele alignment, and Steiger testing assessed the exposure–outcome direction.[27] With one instrument per gene, the Wald-ratio significance restates the underlying sQTL association, and a germline variant acting through splicing is expected to favor an sQTL → MDD Steiger direction. We therefore treated MR and Steiger as directional-consistency checks rather than independent causal evidence.

Cross-disorder colocalization and classification

The best splicing event for each high-confidence gene was colocalized with SCZ, BD, ADHD, PTSD, ANX, AD and PD GWAS. PP.H4 > 0.5 defined sharing. Genes were classified as MDD-only (colocalizing only with MDD among the tested disorders), cross-psychiatric, or broad-CNS/neurodegeneration-related. Because loci were ascertained by MDD significance and GWAS power differs by disorder, these classes indicate relative sharing only.

Complex-LD quality control

The 17q21.31 inversion region was excluded from the main results because its long-range LD and structural variation can produce uninterpretable colocalization. Other complex-LD loci were retained but flagged.

Druggability annotation

Open Targets tractability annotations were used to classify small-molecule and antibody tractability.[28] Targets in the “approved drug” bucket were checked individually, because that bucket may reflect target-family rather than gene-specific approved drugs. Druggability was used for prioritization, not as evidence of isoform-specific intervention.

Colocalization robustness analysis

Robustness was evaluated in two ways. First, coloc was rerun with a conservative shared prior ($p12 = 1\times10^{-6}$) to define a prior-robust core. Second, MDD2025 was LD-clumped with 1000 Genomes European LD ($r^2 < 0.1$, 500 kb),[29,30] and the number of independent MDD association signals within ± 1 Mb of each high-confidence gene was counted. LD-clumping identifies independent GWAS signals but does not establish a single causal variant on either the GWAS or the sQTL side; we therefore used it only to flag loci where the single-causal-variant assumption of coloc.abf is more likely to be violated (multi-signal loci), which would warrant conditional or multi-causal fine-mapping.[31,32]

Sequence-model prediction of splicing effects (AlphaGenome)

For each high-confidence gene we scored its top-associated sQTL variant set, defined as the 15 sQTL variants with the largest |Z| for the best-colocalizing event (a heuristic proxy, not a PIP-based credible

set), with AlphaGenome.[33] We retained the maximum predicted splicing-disruption effect per variant and summarized each gene by its per-gene maximum. A background gene set was built with the same 15-variant construction and scored identically. Group differences were assessed by a one-sided Mann-Whitney test of the per-gene maximum, by the proportion of genes exceeding $\Delta > 0.2$ and $\Delta > 0.5$ (Fisher's exact test), and by a leave-one-gene-out check. Because the per-gene maximum can rise with the number of variants returning a finite score, and the two sets are not matched on this count, we treated the comparison as exploratory.

Gene-structural localization of splicing events

For priority genes, the dominant splicing cluster was mapped onto normalized GENCODE transcript structure and classified as CDS, 5'UTR/alternative-first-exon, or other non-coding localization. This annotation is descriptive and does not establish the exact isoform or protein product.

Cell-type expression context (descriptive)

To indicate where priority splicing-effector genes are expressed, we used the GSE213982 human dorsolateral-prefrontal-cortex single-nucleus RNA-seq dataset. Nuclei were grouped into major cortical cell classes: excitatory neurons (ExN), inhibitory neurons (InN), astrocytes (Ast), oligodendrocyte precursors (OPC), oligodendrocytes (Oli) and microglia (Mic). For each priority gene and cell class, we computed mean log1p-normalized expression and z-scored values across cell classes within each gene. The resulting gene-wise relative expression values were clipped to ± 2 and displayed as a heatmap. This describes expression distribution only and does not resolve the cell-type origin of splicing or test cell-type-specific genetic risk.

Statistics and reproducibility

Analyses used Python, R and command-line tools. Colocalization used a Python implementation of the approximate Bayes factor (ABF) framework, validated against the coloc R package (coloc.abf) on representative loci; LD clumping used plink2; and figures were generated in R. Fisher's exact test compared proportions unless otherwise specified. Main thresholds were PP.H4 > 0.8 for high confidence, PP.H4 > 0.5 for extended or sharing analyses, $p12 = 1\times10^{-6}$ for conservative prior sensitivity, and AlphaGenome $\Delta > 0.2$ for predicted splicing-disruption effects.

Results

A brain splicing-effector gene map for MDD (56-gene high-confidence set)

Across 218 MDD2025 risk regions, 56 brain splicing-effector genes reached PP.H4 > 0.8. Thirty-five exceeded PP.H4 > 0.9, with a median maxPP.H4 of 0.92 and a maximum of 0.999. The extended PP.H4 > 0.5 set contained 132 genes before mapping cleanup and 17q21.31 exclusion (extended-set median 0.74), compared with ~ 22 genes recovered using the earlier 2018 GWAS. Several genes were supported

by multiple splicing events, including GTF2IRD2B (6 events) and RERE and CTC1 (5 events each). The high-confidence genes were distributed across chromosomes rather than concentrated in a single region (Fig. 1b). The discovery workflow and evidence funnel are summarized in Fig. 1a,e.

Most splicing-effector genes show no expression-level colocalization

Using matched BigBrain eQTLs, only 4/56 high-confidence genes showed eQTL-MDD colocalization at PP.H4 > 0.5, and 3/56 at PP.H4 > 0.8. Thus 52/56 genes (93%) were not captured by expression-eQTL colocalization at the lenient threshold, and 53/56 (95%) at the stringent threshold. We report the more conservative 93% as the primary value. MetaBrain provided a concordant cross-resource check: among the 43 high-confidence genes with MetaBrain cis-eQTL data, 37/43 (86%) at the lenient threshold and 39/43 (91%) at the stringent threshold were likewise not captured by expression-eQTL colocalization. A similar 93% of the BigBrain extended set was not captured, indicating that this pattern was not driven by a single eQTL dataset.

We then separated orthogonality from limited eQTL power. Twenty-nine genes (52%) had a detectable cis-eQTL that did not colocalize with MDD. These genes formed a high-confidence expression-orthogonal core that includes CACNA1C, CELF1, MKRN1, NDUFS3, PBRM1 and IP6K1. The remaining 23 genes lacked a detectable eQTL, so their non-colocalization is power-limited. We use “not captured by eQTL colocalization” for the full 52/56 set and reserve “orthogonal core” for the 29 well-supported genes (Fig. 1d).

Figure 2 shows regional colocalization at RERE, EIF4E3, CACNA1C and CELF1. At each locus, the MDD GWAS and brain sQTL peaks aligned near the same lead variant, whereas the eQTL signal was absent or peaked elsewhere.

Directional consistency of splicing MR

Of the 56 high-confidence genes, 45 had a usable single-instrument sQTL; for all 45 the Steiger direction was sQTL → MDD, and 42 reached nominal Wald-ratio significance (per-gene estimates in Supplementary Table 2). Because each estimate rests on a single lead sQTL, this significance restates the underlying sQTL association rather than adding independent evidence, and a uniform sQTL → MDD Steiger direction is expected for germline variants acting through splicing; we therefore read MR and Steiger as directional-consistency checks, not causal proof (Fig. 3a; colocalization robustness in Fig. 3b,c).

Cross-disorder colocalization indicates sharing across psychiatric disorders

Among 56 high-confidence genes with cross-disorder data, 20 colocalized only with MDD, 35 with at least one other psychiatric disorder, and 1 with AD/PD. In the extended set of 132 testable genes, only 1 colocalized with AD/PD. Cross-psychiatric genes included the calcium-channel gene CACNA1C (MDD = 0.93, SCZ = 0.82, BD = 0.84), the transcription factor CREB1, the lipid-metabolism and myelin genes FADS2 and MYRF, and the splicing-related candidates CELF1 and RERE. Within MDD risk loci, high-confidence splicing effects were shared more often with other psychiatric disorders than with AD/PD (Fig. 4; Supplementary Table 3). Because loci were defined by MDD significance, AD/PD signals at these loci are inherently weaker. This is therefore a relative, exploratory classification rather than evidence of absolute disease specificity (see Discussion).

Selected candidate genes

Several high-confidence genes have biological or translational relevance. CACNA1C encodes a voltage-gated calcium-channel subunit, shows splicing-level colocalization in MDD, SCZ and BD, and is a recurrent cross-psychiatric risk gene.[34,35] Its human-brain alternative-splicing profile is highly complex and may contribute to this association.[36] CELF1 links the map to RNA-processing networks, CREB1 to mood regulation and plasticity, and FADS2 and MYRF to lipid metabolism and myelin.[37,38] RERE, CTC1 and GTF2IRD2B were supported by multiple splicing events and are priority candidates for functional study. Multi-evidence prioritization of the top-ranked candidates is summarized in Fig. 5.

Druggability annotation

Open Targets tractability tiers classified all 56 high-confidence genes. Seventeen were small-molecule tractable. Four genes (CACNA1C, HDAC3, NDUFS3, RPS15A) fell in the “approved drug” bucket, but individual verification showed a directly relevant approved class only for CACNA1C (calcium-channel blockers) and HDAC3, which is engaged by approved non-selective HDAC inhibitors rather than an HDAC3-selective drug. NDUFS3 and RPS15A had no relevant approved drug. Nine additional genes were antibody-tractable. These annotations reflect target-level development potential, not isoform-selective therapeutic feasibility (Fig. 5b,c; Supplementary Table 4). As descriptive context, single-nucleus RNA-seq (GSE213982) showed that priority genes were expressed predominantly in excitatory and inhibitory neurons, with cell-type-consistent exceptions such as MYRF in oligodendrocytes and FADS2 in astrocytes (Fig. 5d).

Colocalization robustness: prior sensitivity and multi-signal loci

Under conservative p12 = 1×10^{-6}, 13/56 genes remained above PP.H4 > 0.8, defining a prior-robust core (all PP.H4 $\gtrsim$ 0.95; full list in Supplementary Table 2). This core retained EIF4E3, MKRN1, PPP2R5B, CCNC, RERE and AKAP6, among others. Some moderate-strength candidates (CACNA1C, CELF1, CREB1; PP.H4 $\approx$ 0.93) reached high confidence under the default prior but not the conservative prior. We therefore report the 56-gene default-prior set as the main set and the 13-gene core as a high-priority subset. LD

clumping identified 325 independent MDD2025 signals genome-wide. Within ± 1 Mb of high-confidence genes, 38/56 loci (68%) contained no additional independent MDD association signal beyond the lead. LD-clumping identifies independent GWAS signals but does not by itself prove a single causal variant; the 18 multi-signal loci (2–4 signals), including CELF1, EIF4E3, FADS2, PBRM1 and MYRF, are the most likely to violate the coloc.abf single-causal-variant assumption and are flagged for future conditional or multi-signal analyses (Fig. 3b,c). The gene-level maximum-PP.H4 rule did not bias selection toward multi-event genes: maximum PP.H4 was uncorrelated with the number of splicing events (Spearman ρ = −0.13, P = 0.33), 45% of high-confidence genes were single-event, and 45 of the 56 also passed a stricter median-event PP.H4 > 0.8 rule.

Gene-structural localization of splicing events

The dominant sQTL clusters of the 12 named prior-robust genes and 3 cross-psychiatric candidates (CACNA1C, CELF1, CREB1) were mapped onto gene structure. Among 15 annotatable genes, 6 dominant events lay in CDS regions and 9 in 5'UTR or alternative-first-exon regions (per-gene assignments in Fig. 6a). CDS events are priority candidates for protein-coding consequences, whereas 5'UTR/alternative-first-exon events may affect transcript 5'-end structure or translational regulation. This gene-structural classification does not establish isoform or protein consequences.

Top-associated sQTL variants show suggestive sequence-model support for splicing effects

Colocalization does not show that the shared variant directly alters splicing. We therefore scored each gene's top-associated sQTL variant set (the 15 top-|Z| sQTL variants of its best-colocalizing event) with AlphaGenome and compared the per-gene maximum predicted splicing-disruption effect against a comparison background gene set.

Splicing-effector genes showed higher predicted disruption than background genes (median per-gene maximum Δ = 0.29 vs 0.18). The one-sided Mann-Whitney test was significant (P = 0.028), whereas the two-sided test was not (P = 0.055). More splicing-effector genes crossed the effect thresholds (62% versus 45% at Δ > 0.2, odds ratio 1.99, Fisher P = 0.066; 32% versus 20% at Δ > 0.5), and the direction held under leave-one-gene-out (maximum P = 0.15). Because the variant sets are |Z|-ranked rather than PIP-based, the two groups are not matched on finite-score count, and significance is one-sided only, we read this as exploratory sequence-level support consistent with a splicing mechanism rather than independent evidence; direct consequences require transcript-level and functional validation (Fig. 6b,c; Supplementary Table 5).

Discussion

Main findings

Using MDD2025 and BigBrain, we identified 56 high-confidence brain splicing-effector genes for MDD. Most were not captured by brain eQTL colocalization, and 29 formed a well-supported expression-orthogonal core. MR/Steiger supported directional consistency, 38/56 loci were single-signal regions, and 13 genes remained robust under a conservative coloc prior. Cross-disorder and tractability annotations prioritized candidates including CACNA1C, CELF1, CREB1, FADS2, MYRF and HDAC3. AlphaGenome provided exploratory sequence-level evidence consistent with predicted effects on splicing.

Relationship to existing work

Previous disease-genetic and brain-sQTL studies established that genetic effects on splicing can be partly independent of gene expression,[6,12] and BigBrain demonstrated this systematically across several neurological and psychiatric disorders.[11] Building on this framework, our study provides a systematic MDD splicing-effector resource. The analysis combines MDD2025, BigBrain sQTLs, matched eQTL controls, conservative-prior and signal-complexity checks, cross-disorder annotation, druggability and sequence-effect prediction. BigBrain's original pan-disease analysis included MDD, but used an earlier, smaller GWAS and did not provide an MDD-focused synthesis with matched eQTL control, cross-disorder sharing and tractability annotation.

The most directly related study is Goes et al. (2025).[14] In 846 MDD brain samples from subgenual cingulate and amygdala, that study identified 89 splicing-QTL genes at MDD risk loci, including 48 captured only by splicing, and reported synaptic genes such as NRXN1, GABRA1, RBFOX and CACNA1C. Our results are directionally concordant. The main differences are scale and annotation depth. We use BigBrain (10 725 brain RNA-seq samples, multi-region mmQTL meta-analysis) and MDD2025 (635 genome-wide significant loci), and run BigBrain eQTLs in parallel using the same regions, GWAS and colocalization pipeline. This matched design identifies a 29-gene subset whose MDD associations are captured by splicing but not by same-locus eQTL colocalization and adds a conservative prior-robust core, psychiatric/neurodegenerative cross-disorder controls, druggability and AlphaGenome sequence-effect prediction. The two studies therefore provide complementary evidence.

This work also differs from OPERA, which treats splicing as one molecular layer among several and draws its molecular QTLs mainly from GTEx and PsychENCODE.[16,39,40] Here, splicing is the primary phenotype, BigBrain brain sQTLs are the resource, and same-locus expression is the matched control. The result is a focused map of genes whose MDD association is better captured at the splicing level than by current bulk-brain eQTL colocalization.

Implications for the genetic interpretation of MDD

The map shows that expression-based analysis alone may miss part of the MDD risk-gene set. sQTLs capture RNA-processing changes that may affect exon choice, protein domains, RNA stability or

transcript usage.[8–10] The key result is not the absence of expression regulation, but lack of colocalization in current bulk-brain eQTL resources. We therefore interpret the 93–95% proportion as "not captured by eQTL colocalization," with the 29-gene subset providing the strongest expression-orthogonal evidence. Such genes (which carry a splicing event consistent with sharing a causal variant with the MDD GWAS but no colocalizing eQTL) can be overlooked in an eQTL/TWAS framework and are recovered by sQTL colocalization.

Cross-disorder sharing and candidate targets

Cross-disorder results place many candidates within the psychiatric spectrum, consistent with the shared genetic architecture and pleiotropy documented across psychiatric disorders.[41–43] CACNA1C is the clearest example, with MDD, SCZ and BD sharing and an established drug-target background. CELF1, CREB1, FADS2 and MYRF extend the map toward RNA processing, plasticity, lipid metabolism and myelin. Druggability remains target-level information. Existing calcium-channel blockers or HDAC inhibitors act on the target protein, not selectively on the implicated risk splicing isoforms. Splicing-mediated risk may instead require isoform-selective strategies, such as antisense oligonucleotides[44] or splicing-modulating small molecules. These annotations therefore prioritize follow-up targets rather than imply ready therapeutic repurposing. Establishing more symmetric disease specificity will require designs anchored on each disorder's own significant loci.

Methodological interpretation and alternative explanations

The main technical risk is signal attribution: coloc.abf assumes one causal variant per trait per region. Our signal census found no additional independent MDD signal at 38/56 loci, consistent with this assumption on the GWAS side, though it cannot establish a single causal variant on either trait; the 18 multi-signal loci need re-examination with in-sample LD, conditional analysis or multi-causal fine-mapping. We attempted coloc.susie with 1000 Genomes reference LD, but LD mismatch between the external reference panel and the MDD2025/BigBrain summary statistics prevented stable convergence, so we did not report per-signal posteriors. MR/Steiger, resting on single lead-sQTL instruments, likewise provides directional rather than independent causal support, in line with appraisals of TWAS/colocalization methods.[45]

Limitations

Several further limitations apply. The map is anchored on genome-wide significant MDD loci and does not capture sub-threshold splicing effects. Evidence comes from public summary statistics without functional perturbation. BigBrain is bulk brain, so the cell-type origin of the splicing effects cannot be resolved, and large-scale brain cell-type sQTL resources are not yet available. The single-nucleus expression heatmap in Fig. 5d is descriptive context and neither localizes the splicing signal to a cell type nor implies cell-type-specific genetic risk. A formal single-nucleus enrichment test was not significant. Lack of eQTL colocalization is relative to current bulk-brain resources, and gene-level maximum PP.H4 collapses event heterogeneity, although a sensitivity analysis indicated that this does

not bias gene selection toward multi-event genes (Results). Cross-disorder classes are relative because loci were MDD-ascertained. Druggability is target-level, not isoform-specific. AlphaGenome support is exploratory (one-sided only and based on |Z|-ranked variant sets) and remains a computational prediction. Accordingly, these findings constitute a genetic resource rather than a functionally validated causal model.

Conclusion

Integrating MDD2025 with BigBrain brain sQTLs identified 56 high-confidence MDD splicing-effector genes. Most were not captured by brain eQTL colocalization, and a 29-gene subset had a significant same-locus cis-eQTL that did not colocalize with MDD. The map nominates translational candidates including CACNA1C, HDAC3, CELF1, CREB1, RERE, FADS2 and MYRF. It provides a complementary regulatory layer for MDD genetic-risk interpretation and functional follow-up.

Declarations

Competing interests:

The authors declare no competing interests.

Conflict of Interest

The authors declare no competing interests.

Ethics

This study is a secondary analysis of publicly available, de-identified GWAS and brain-QTL summary statistics and did not involve new human-subjects data collection. Ethical approval and informed consent were obtained by the original studies. No additional ethical approval was required.

Funding

This work was supported by the Talent Development Program of Zhejiang Chinese Medical University (grant no. 2024RCZXZK14). The funder had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

Author Contributions

The authors' contributions are described using the CRediT taxonomy. **Bozhi Li**: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Project administration. **Weijia Gu**: Formal analysis, Validation, Writing – review & editing. **Xiaqiu Wu**: Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgements

We thank the participants and investigators of the studies whose summary-level data made this work possible. This research used data generated by the Psychiatric Genomics Consortium (PGC) and the Adams et al. major depressive disorder GWAS, the BigBrain brain-QTL resource, the MetaBrain consortium, the Bellenguez et al. and Nalls et al. neurodegenerative-disease GWAS efforts, the 1000 Genomes Project, and the Open Targets Platform; we are grateful to these consortia for making their data publicly available.

Data Availability

All data used are public. The MDD GWAS is PGC/Adams 2025 (MDD2025). The comparison GWAS (SCZ, BD, ADHD, PTSD, ANX, AD, PD) are from PGC, the GWAS Catalog and figshare. Brain splicing-QTLs and the primary control brain expression-eQTLs are from the BigBrain resource. As of this analysis, BigBrain is a preprint (medRxiv 2025.09.25; doi:10.1101/2025.09.25.25336663), with data archived at Zenodo: cis-sQTL doi:10.5281/zenodo.17153730 and cis-eQTL doi:10.5281/zenodo.17226890 (both accessed in 2025). MetaBrain brain eQTLs were used for independent cross-resource validation. The single-nucleus cell-type expression context (Fig. 5d) used the public human dorsolateral-prefrontal-cortex snRNA-seq dataset GSE213982. Sequence-level prediction used the AlphaGenome free non-commercial API (DeepMind) for variant-effect scoring only, without modifying model weights. All analysis code and derived statistical tables (colocalization results, orthogonality tiers, MR/Steiger, cross-disorder matrix, single-signal census, druggability and AlphaGenome scores) are deposited at Zenodo (DOI: 10.5281/zenodo.21352283). The record will be made publicly available upon publication.

Code availability

All custom code used to generate the results, tables and figures is available at the same Zenodo deposit as the data (DOI: 10.5281/zenodo.21352283), released under the MIT license. This includes the colocalization pipeline (approximate-Bayes-factor implementation, validated against the coloc R package), the matched brain-eQTL control, splicing Mendelian randomization and Steiger testing, cross-disorder colocalization and classification, the LD-based single-signal census, Open Targets druggability annotation, AlphaGenome variant scoring, and all figure-generation scripts. Analyses used Python 3 (NumPy, pandas, SciPy), R 4.x (coloc, susieR, ggplot2), PLINK 2.0 and tabix/htslib; exact package versions and run instructions are provided in the repository README.

References

1. Adams M.J. et al. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. Cell 188, 640–652 (2025).
2. Howard D.M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. 22, 343–352 (2019).
3. Giambartolomei C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).
4. Zhu Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487 (2016).
5. Gusev A. et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. 48, 245–252 (2016).
6. Li Y.I. et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016).
7. Park E. et al. The expanding landscape of alternative splicing variation in human populations. Am. J. Hum. Genet. 102, 11–26 (2018).
8. Dredge B.K., Polydorides A.D. & Darnell R.B. The splice of life: alternative splicing and neurological disease. Nat. Rev. Neurosci. 2, 43–50 (2001).
9. Vuong C.K., Black D.L. & Zheng S. The neurogenetics of alternative splicing. Nat. Rev. Neurosci. 17, 265–281 (2016).
10. Scotti M.M. & Swanson M.S. RNA mis-splicing in disease. Nat. Rev. Genet. 17, 19–32 (2016).
11. Réal A. et al. Mapping genetic effects on splicing in ten thousand post-mortem brain samples reveals novel mediators of neurological disease risk (BigBrain). medRxiv https://doi.org/10.1101/2025.09.25.25336663 (2025). (Preprint at submission; data: Zenodo 10.5281/zenodo.17153730 [sQTL], 17226890 [eQTL].)
12. Takata A., Matsumoto N. & Kato T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. Nat. Commun. 8, 14519 (2017).
13. Walker R.L. et al. Genetic control of expression and splicing in developing human brain informs disease mechanisms. Cell 179, 750–771 (2019).
14. Goes F.S. et al. Large-scale transcriptomic analyses of major depressive disorder reveal convergent dysregulation of synaptic pathways in excitatory neurons. Nat. Commun. 16, 3981 (2025).
15. Wray N.R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681 (2018).
16. Nisbet L. et al. Integrating multi-omics summary data identifies candidate molecular mechanisms for major depression (OPERA). Biol. Psychiatry 99, 1120–1129 (2026). doi:10.1016/j.biopsych.2025.11.024.
17. Trubetskoy V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604, 502–508 (2022).

18. Mullins N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat. Genet. 53, 817–829 (2021).
19. Demontis D. et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. Nat. Genet. 55, 198–208 (2023).
20. Nievergelt C.M. et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. Nat. Commun. 10, 4558 (2019). (PGC-PTSD freeze 3, European-ancestry summary statistics.)
21. Strom N.I. et al. Genome-wide association study of major anxiety disorders in 122 341 European-ancestry cases identifies 58 loci and highlights GABAergic signaling. Nat. Genet. 58, 275–288 (2026).
22. Bellenguez C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. 54, 412–436 (2022).
23. Nalls M.A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 18, 1091–1102 (2019).
24. Li Y.I. et al. Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet. 50, 151–158 (2018).
25. Zeng B. et al. Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits (mmQTL). Nat. Genet. 54, 161–169 (2022).
26. de Klein N. et al. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases (MetaBrain). Nat. Genet. 55, 377–388 (2023).
27. Hemani G., Tilling K. & Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 13, e1007081 (2017).
28. Ochoa D. et al. Open Targets Platform: supporting systematic drug–target identification and prioritisation. Nucleic Acids Res. 49, D1302–D1310 (2021).
29. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526, 68–74 (2015).
30. Chang C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7 (2015).
31. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. 17, e1009440 (2021).
32. Zou Y., Carbonetto P., Wang G. & Stephens M. Fine-mapping from summary data with the "Sum of Single Effects" model (SuSiE-RSS). PLoS Genet. 18, e1010299 (2022).
33. Avsec Ž. et al. Advancing regulatory variant effect prediction with AlphaGenome. Nature 649, 1206–1218 (2026). doi:10.1038/s41586-025-10014-0.
34. Moon A.L. et al. CACNA1C: association with psychiatric disorders, behavior, and neurogenesis. Schizophr. Bull. 44, 958–965 (2018).

35. Kabir Z.D., Lee A.S. & Rajadhyaksha A.M. L-type Ca^{2+} channels in mood, cognition and addiction. J. Physiol. 594, 5823–5837 (2016).
36. Clark M.B. et al. Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene CACNA1C in human brain. Mol. Psychiatry 25, 37–47 (2020).
37. Wu X. et al. Brain tissue- and cell type-specific eQTL Mendelian randomization reveals efficacy of FADS1 and FADS2 on cognitive function. Transl. Psychiatry 14, 77 (2024).
38. Stacey D. et al. Integrative genomic characterization of the FADS1-2-3 mood disorder risk locus: implications for the role of polyunsaturated fatty acids. J. Affect. Disord. 412, 122140 (2026). doi:10.1016/j.jad.2026.122140.
39. Gandal M.J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder (PsychENCODE). Science 362, eaat8127 (2018).
40. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318–1330 (2020).
41. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 179, 1469–1482 (2019).
42. Sullivan P.F. & Geschwind D.H. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. Cell 177, 162–183 (2019).
43. Watanabe K. et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. 51, 1339–1348 (2019).
44. Finkel R.S. et al. Nusinersen versus sham control in infantile-onset spinal muscular atrophy. N. Engl. J. Med. 377, 1723–1732 (2017).
45. Wainberg M. et al. Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. 51, 592–599 (2019).

Figures

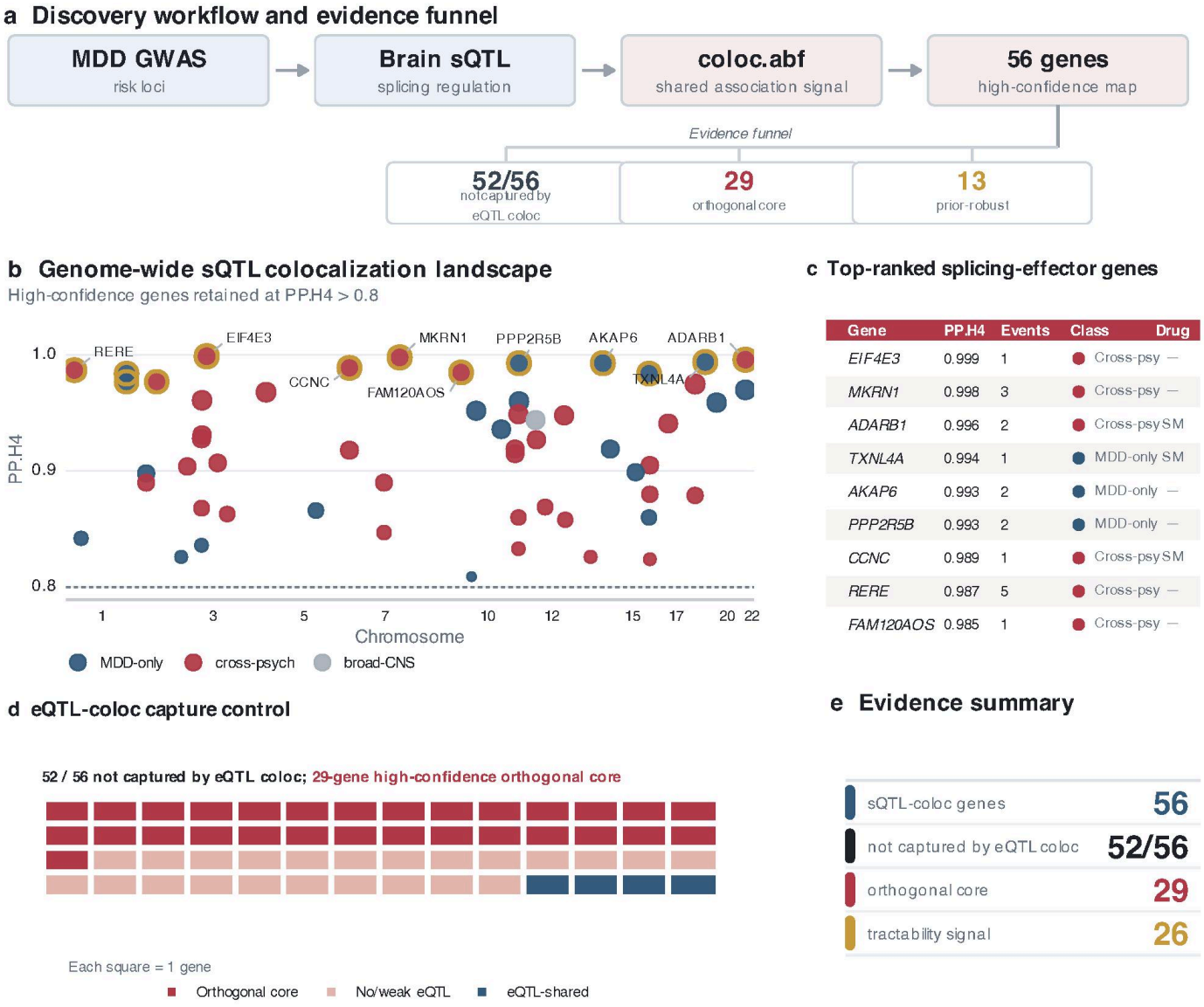


Figure 1

Overview of the MDD brain splicing-effector gene map. (a) Discovery workflow and evidence funnel: MDD GWAS risk loci → brain sQTL → coloc.abf test for a shared association signal → 56 high-confidence splicing-effector genes, including 52/56 not captured by eQTL colocalization, a 29-gene orthogonal core, and a 13-gene prior-robust core. (b) Genome-wide sQTL colocalization landscape of the 56 high-confidence genes (x = chromosome, y = PP.H4; point color = cross-disorder class: blue = MDD-only, red = cross-psychiatric, gray = broad-CNS; gold ring = conservative-prior robust gene; dashed line, PP.H4 = 0.8 threshold). (c) Top-ranked splicing-effector genes ranked by PP.H4 (number of splicing events, cross-disorder class, and druggability tier; SM = small-molecule, Ab = antibody, limited = no tractability signal). (d) eQTL colocalization control, each square = one gene: 52/56 genes are not captured by expression-eQTL colocalization, of which 29 form a high-confidence orthogonal core (dark red = orthogonal core, a significant cis-eQTL that does not colocalize with MDD; light red = no/weak eQTL, orthogonality untested; blue = shared with an expression eQTL). (e) Evidence summary: 56 sQTL-colocalizing genes,

52/56 not captured by eQTL colocalization, a 29-gene orthogonal core, and 26 genes carrying a druggability (tractability) signal.



Figure 2

Regional colocalization (locus-zoom) at four representative splicing-effector loci. **a**–**d**, RERE (PP.H4 = 0.987), EIF4E3 (PP.H4 = 0.999), CACNA1C (PP.H4 = 0.927) and CELF1 (PP.H4 = 0.919). For each gene,

three vertically stacked tracks show the $-\log_{10}$P of the MDD2025 GWAS (top), the brain splicing QTL (middle) and the brain expression eQTL (bottom) against chromosomal position. Variants are binned and colored by linkage disequilibrium to the lead splicing-QTL variant (LD r^2, 1000 Genomes European samples; blue to red for increasing r^2), and the lead variant is shown as a purple diamond; the dashed line in the GWAS track marks genome-wide significance ($P = 5\times10^{-8}$). At each locus the GWAS and splicing-QTL peaks coincide at the same high-LD lead variant, whereas the expression eQTL either peaks at a different variant (EIF4E3, CACNA1C, CELF1) or is undetectable (RERE has no detectable BigBrain cis-eQTL; orthogonality is therefore not evaluable for this locus), illustrating that these MDD risk signals colocalize at the splicing but not the expression level. Colocalization uses coloc.abf. The underlying GWAS, sQTL and eQTL association statistics for these loci are available in the data repository (Data Availability).

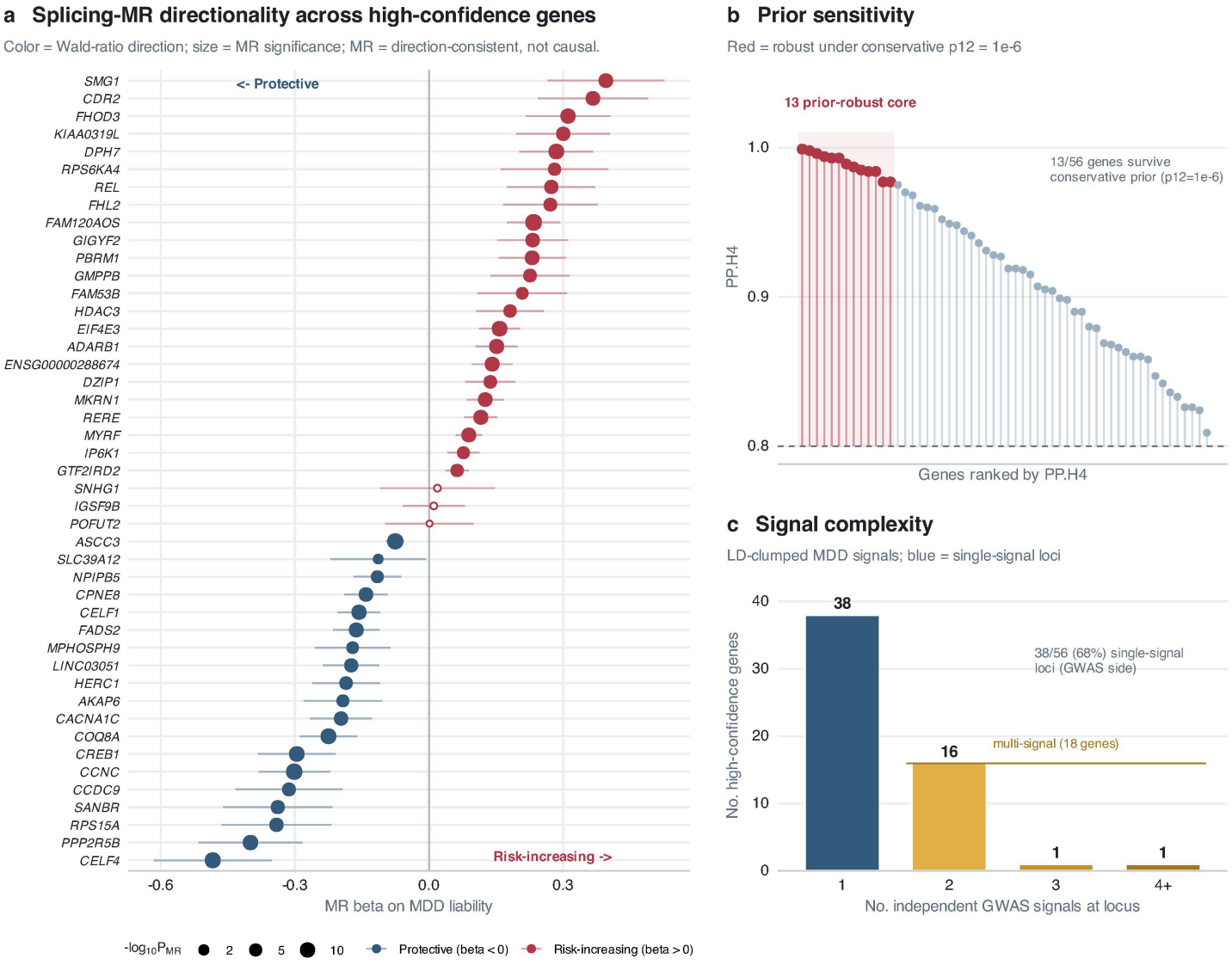


Figure 3

Directional consistency and colocalization robustness. (a) Splicing MR for the 45 high-confidence genes with a usable single-instrument sQTL (Wald-ratio β and 95% confidence interval); all 45 had a cis-sQTL→MDD Steiger direction, and given the single-instrument design, MR describes directional consistency rather than independent causality. (b) Prior sensitivity: the 56 high-confidence genes ranked by PP.H4, with the prior-robust core (still > 0.8 under the conservative prior p12 = 1×10^{-6}) in red. (c) Distribution of the number of independent MDD signals within each high-confidence locus (±1 Mb; 1000 Genomes LD-clumping, $r^2 < 0.1$): 38/56 (68%) are single-signal loci (no additional independent MDD association signal detected), while the remaining 18 are multi-signal loci (2–4 signals) where the coloc.abf single-causal-variant assumption is more likely to be violated.

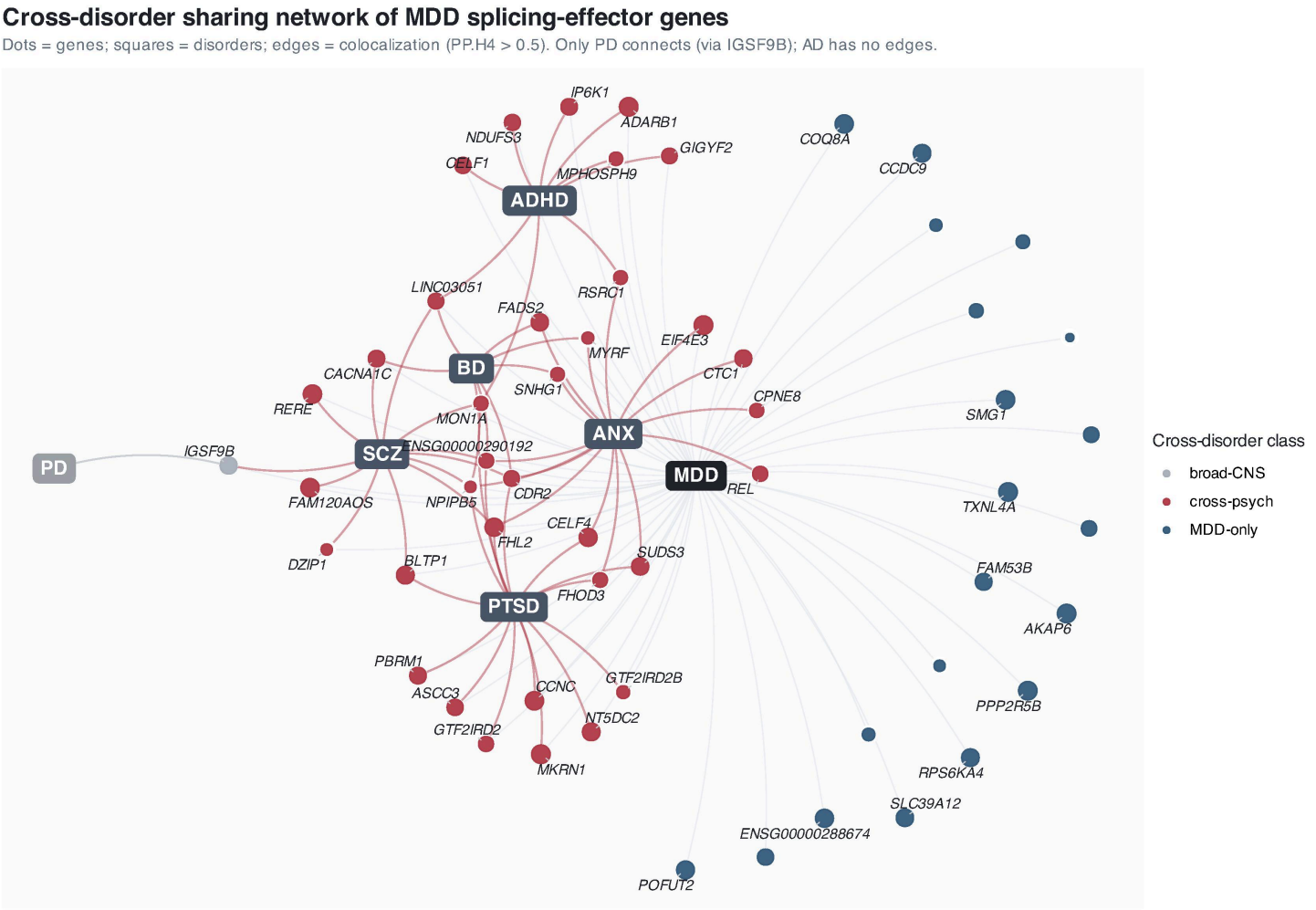


Figure 4

Cross-disorder sharing network of splicing-effector genes (force-directed layout). Squares are disorder nodes and dots are high-confidence splicing-effector genes (color indicates cross-disorder class: blue = MDD-only, red = cross-psychiatric, gray = broad-CNS; size indicates PP.H4); edges indicate that a gene colocalizes in the corresponding disorder (PP.H4 > 0.5). Cross-psychiatric genes (red) connect multiple disorder nodes (SCZ, BD, ADHD, PTSD, ANX) and lie toward the network center, MDD-only genes (blue)

connect only to MDD and lie at the periphery; among the neurodegenerative contrasts, only the PD node has a connection (a single gene, IGSF9B), whereas AD had no colocalizing genes and is therefore not displayed. The quantitative PP.H4 matrix is in Supplementary Table 3.



Figure 5

Evidence, tractability and cellular expression context of candidate splicing-effector genes. (a) Candidate-gene evidence landscape: x-axis, sQTL-MDD colocalization PP.H4; y-axis, expression-orthogonality tier (orthogonal core, no/weak eQTL, or eQTL-shared); bubble size, number of supporting splicing events; color, cross-disorder class (blue = MDD-only, red = cross-psychiatric, gray = broad-CNS); gold rings mark the prior-robust core. Genes with high PP.H4 in the orthogonal-core band are the strongest splicing-specific candidates. (b) Open Targets druggability tiers across the 56 high-confidence genes. (c) Small-molecule-tractable candidate targets with their associated drug classes or mechanisms and cross-disorder class: CACNA1C is engaged by approved calcium-channel blockers (e.g. nifedipine) and HDAC3 by approved non-selective HDAC inhibitors (e.g. panobinostat); NDUFS3 (mitochondrial complex I inhibitors, preclinical) and RPS15A (ribosomal protein) are tractable but lack a specific approved drug. (d) Single-nucleus expression heatmap for priority genes from the splicing map (single-nucleus RNA-seq, GSE213982); tile color indicates gene-wise relative mean log1p expression across major cortical cell classes (ExN, InN, Ast, OPC, Oli, Mic). This panel is descriptive expression context and does not localize the splicing signal to a cell type. CACNA1C, with both cross-psychiatric sharing and approved calcium-channel drugs, is a priority candidate.

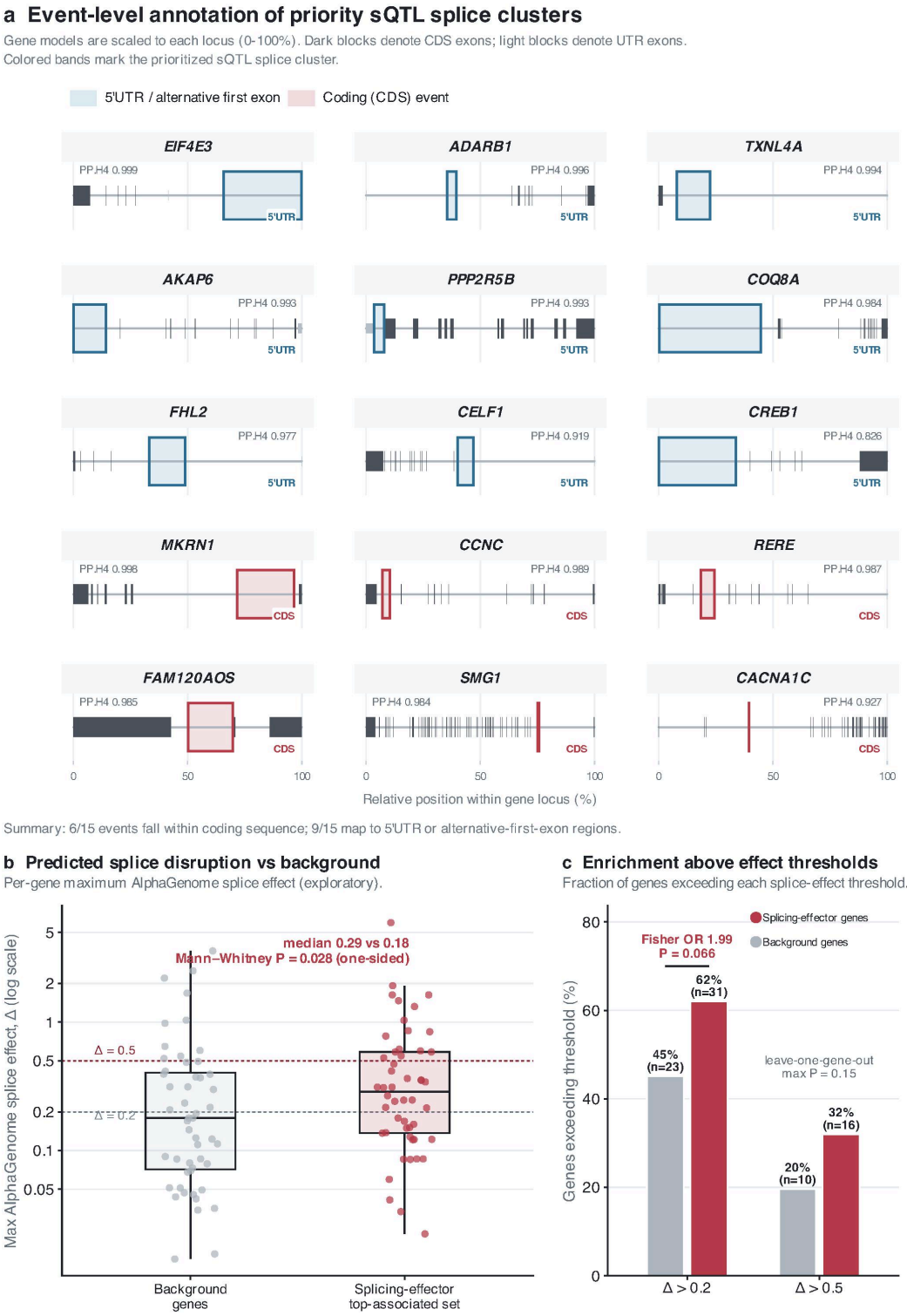


Figure 6

Event-level structure and sequence-level effect prediction of priority splicing candidates. (a) Gene-structural localization of the dominant sQTL splicing events: each panel shows one gene's transcript model (normalized to 0–100% across the locus; dark blocks, CDS exons; light blocks, UTR exons), with colored bands marking the dominant sQTL intron cluster (blue, 5'UTR/alternative-first-exon events; red, CDS events) and the gene's colocalization PP.H4. Among 15 annotatable genes, 6 dominant events lie in

the CDS and 9 in 5'UTR/alternative-first-exon regions; the classification is based on gene-structural localization and does not represent a validated transcript or protein change. (b) Per-gene maximum AlphaGenome predicted splicing-disruption effect (Δ) for splicing-effector genes (scored over each gene's top-associated sQTL variant set of 15 top-|Z| sQTL variants) versus a background gene set built the same way; each point is one gene. Splicing-effector genes show a higher median per-gene maximum Δ (0.29 vs 0.18) under a one-sided Mann-Whitney test (one-sided P = 0.028; the two-sided test was not significant). (c) Proportion of genes exceeding the predicted-effect thresholds: at Δ > 0.2, 62% of splicing-effector genes versus 45% of background (Fisher odds ratio = 1.99, P = 0.066); at Δ > 0.5, 32% versus 20%. The direction is not driven by any single gene (leave-one-gene-out maximum P = 0.15). The analysis is exploratory (|Z|-ranked variant sets, one-sided significance) and is a computational prediction, not experimental validation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.xlsx